

5 **CANCER MONITORING BY ABERRANT PROMOTOR METHYLATION OF  
THE TRANSCRIPTION FACTOR GENES PAX5 ALPHA PAX5 BETA, NOVEL  
LOOP HELIX LOOP PROTEIN, NOVEL GENE 2, AND BETA3 GENES**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

10 This application claims the benefit of the filing of United States of America  
provisional patent application Serial No. 60/348,407, entitled "Inactivation of PAX5  
Alpha and Beta Genes in Cancer," filed on October 18, 2001, and the specification thereof  
is incorporated herein by reference.

**GOVERNMENT RIGHTS**

15 The U.S. Government has certain rights this invention and the right in limited  
circumstances to require the patent owner to license others on reasonable terms as  
provided for by the terms of Contract No. DAMD17-99-1-9258 awarded by U.S.  
Department of Defense.

20 **BACKGROUND OF THE INVENTION**

**Field of the Invention (Technical Field):**

The present invention relates to the field of cancer screening, diagnosis and  
treatment monitoring, specifically the inactivation of PAX5 alpha and beta genes, and  
other identified genes, by promoter hypermethylation marking.

25 **Background Art:**

30 The prognosis for patients with cancer is primarily dependent on the stage of the  
tumor at the time of clinical diagnosis. Human neoplasms are known to arise through a  
progressive accumulation of genetic alterations in protooncogenes and tumor suppressor  
genes. Because these genetic changes are retained as a tumor grows, they can serve as  
markers for detection of cancer. Therefore, it is useful that these genetic markers be  
identified, so they can be used for early detection, to monitor therapy, and to predict  
outcome of the disease.

Drs. James Herman and Stephen Baylin developed a technique known as methylation specific polymerase chain reaction (MSP) to detect gene-specific alterations in promoter methylation within tumor cells. This technique, described in U.S. Patent Nos. 5,786,146 and 6,017,704, has been used to detect promoter hypermethylation of cancer genes (e.g., p16, Rb ER, and MGMT) and can detect one copy of the methylated gene in a background of 1,000 unmethylated copies.

There is now evidence linking dysregulation of the DNA methylation machinery with tumorigenesis and tumor progression. In neoplasia, DNA methylation patterns increase within the promoter of many tumor suppressor genes. This alteration leads to gene silencing, and serves as an alternative to coding region mutations. Recently, a technique called methylated CpG island amplification was developed as a genome screening approach to identify genes that are differentially methylated in cancer cells. See Patent Cooperation Treaty Application No. PCT/US01/26452, filed August 24, 2001, assigned to the assignee of the present invention, the entirety of which is incorporated herein by reference.

A need remains for an identification of the novel genes inactivated by promoter hypermethylation in cancers, such as human breast cancer, and to determine the commonality for gene inactivation in other solid tumor types. Against this background, the present invention was developed.

#### SUMMARY OF THE INVENTION (DISCLOSURE OF THE INVENTION)

PAX5  $\alpha$  and  $\beta$  genes, and others identified, as markers for detecting, monitoring, and diagnosing human cancer. A polymerase chain reaction-based technique of methylated CpG island amplification, followed by representational difference analysis, identifies these genes methylated in human cancer.

Aberrant methylation of 5' CpG islands is a key epigenetic event in many human cancers. Two of CpG islands are identified and mapped to the 5' UTR region of the PAX5  $\alpha$  and  $\beta$  genes. These genes, located on chromosome 9p13, are transcribed from two distinct promoters and form two alternative first exons that are subsequently spliced to the common exons 2-10. The resulting splice variants encode two distinct transcription

factors important in cell differentiation and embryonic development. Examination of the methylation status of each gene using methylation-specific PCR reveals that both genes are hypermethylated in, for example, breast, lung and colon tumors. Analysis of methylated cell lines and tumors by combined bisulfite restriction analysis and bisulfite sequencing reveals dense methylation patterns within each 5'CpG island, strongly correlating with transcriptional silencing. The identified aberrant promoter methylation is a mechanism for dysregulation of the PAX5 and other identified genes according to the present invention.

Objects, advantages and novel features, and further scope of applicability of the present invention will be set forth in part in the detailed description to follow, taken in conjunction with the accompanying drawings, and in part will become apparent to those skilled in the art upon examination of the following, or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated into and form a part of the specification, illustrate several embodiments of the present invention and, together with the description, serve to explain the principles of the invention. The drawings are only for the purpose of illustrating a preferred embodiment of the invention and are not to be construed as limiting the invention. In the drawings:

Fig. 1 is a table showing the frequency of PAX5  $\alpha$  and PAX5  $\beta$  methylation in tumor cell line primary tumors and non-malignant specimens;

Fig. 2 illustrates a representative methylation-specific polymerase chain reaction analysis according to the present invention;

Figs. 3A and 3B illustrate a confirmation, by methylation-specific polymerase chain reaction analysis, of methylation changes indicative of PAX5 gene inactivation, whereby

PAX5  $\alpha$  transcript was abundant or present in certain cell lines (Fig. 3A), while no transcript was present in other cell lines (Fig. 3B);

5 Figs. 4A and 4B are bar graphs summarizing methylation density, as a percent of CpG sites methylated, of PAX5  $\alpha$  and PAX5  $\beta$  genes, respectively, in tumor cell line primary tumors and non-malignant specimens, according to the invention, indicative of the methylation and silencing of the PAX5  $\alpha$  and PAX5  $\beta$  genes; and

10 Fig. 5 is a table showing the frequency of methylation of the Novel Helix Loop Helix Binding Protein, Novel Gene 2, and Beta3 genes in tumor cell lines, primary tumors and non-malignant specimens, according to the present invention.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS (BEST MODES FOR CARRYING OUT THE INVENTION)

15

Breast cancer is the second leading cause of cancer-related death in the United States, and the number of cases is increasing each year. Mortality from this disease could be reduced greatly through an improved understanding of the molecular alterations contributing to cancer initiation and progression. A mechanism in many sporadic cancers, aberrant promoter hypermethylation, is an epigenetic event involving the methylation of normally unmethylated cytosines within the promoter region of genes (Baylin and Herman, 2000; Jones and Laird, 1999). This change in methylation pattern leads to transcriptional silencing and serves as an alternative to coding region mutation (Baylin and Herman, 2000; Jones and Laird, 1999). In breast cancer, aberrant methylation inactivates numerous genes functioning in key cellular pathways (Yang *et al.*, 2001; Zochbauer-Muller *et al.*, 2001; Du *et al.*, 2001; Burbee; *et al.*, 2001; Toyooka *et al.*, 2001).

30 Cancer genome-wide screening approaches are known, for identifying genes inactivated by promoter hypermethylation. These approaches include methylation-sensitive arbitrarily primed PCR (Gonzalzo *et al.*, 1997; Huang *et al.*, 1997), restriction landmark genomic scanning (Costello *et al.*, 2000), CpG microarrays (Yan *et al.*, 2001),

5 methyl-CpG binding domain chromatography (Brock *et al.*, 2001), and methylated CpG island amplification (MCA) coupled with representational difference analysis (RDA) (Toyota *et al.*, 1999a). The MCA/RDA approach has been used to identify several methylated genes involved in colorectal (Toyota *et al.*, 1999b; Toyota *et al.*, 1999c) and  
10 pancreatic cancers (Ueki *et al.*, 2001). MCA/RDA is a PCR/subtraction hybridization based assay that allows for the rapid amplification and selection of densely methylated CpG rich regions ranging in size from 200 bp to 2 kb.

10 The PAX5 gene plays a role in cell differentiation and embryonic development, and is located on chromosome 9p13. This locus is frequently associated with chromosomal translocations and contains two distinct promoters resulting in two alternative 5' exons ( $\alpha$  and  $\beta$ ) that are spliced to common coding sequences of exons 2–  
10. Using amplicons from the breast cancer cell line MCF7 as the tester, and amplicons from normal breast tissue as the driver, the applicants identified the region 5' UTR and  
15 exon 1 of the transcription factor PAX5.

Using a methylated CpG island amplification technique, applicants identified these two genes, PAX5  $\alpha$  and  $\beta$ , that exhibit promoter hypermethylation in common cancers such as breast, colon and lung. These genes, located on chromosome 9p13, are transcribed  
20 from two distinct promoters and result in the formation of two alternative 5' exons spliced to the common exons 2-10. The resulting splice variants encode two distinct transcription factors important in cell differentiation and embryonic development. Recent evidence by others now demonstrates that expression of these genes is deregulated in several tumor types; however, no one has implemented promoter hypermethylation as a means for  
25 silencing these genes.

The present invention exploits the determination that the PAX5  $\alpha$  and  $\beta$  genes are both frequent targets for aberrant methylation in tumor cell lines, as well as primary tumors from breast, lung, and colon. The invention includes the following PAX5  
30 methylation specific polymerase chain reaction (MSP) primer sequences to assay for promoter methylation:

### Stage 1 MSP Primers

<u>Gene</u>	<u>Primers (5' - 3')</u>	<u>Product Size</u>
PAX5 alpha	gggtttgtatatggagatgttatagg (SEQ ID NO:1) Caacatcacaaaatatccccaacac (SEQ ID NO:2)	389 bp
PAX5 beta	agtttgtgggttgttagtaatgg (SEQ ID NO:3) Caaaaaatcccaaccaccaaacc (SEQ ID NO:4)	328 bp

### Stage 2 MSP Primers

<u>Gene</u>	<u>Primers (5' - 3')</u>	<u>Product Size</u>
PAX5 alpha	ataaaagtttggggcggcgc (SEQ ID NO:5) Gcgccccaacgcgcgcg (SEQ ID NO:6)	166 bp
PAX5 beta	gagttgagtttcgggcggc (SEQ ID NO:7) Gccgccgccgcgtcg (SEQ ID NO:8)	124 bp

Using the MSP assay, applicants examined the methylation status of the Pax5  $\alpha$  and  $\beta$  genes in breast, lung, and colon cancers. Pax5  $\alpha$  is methylated in approximately 70% of breast and lung tumors and 13% of colon tumors. The Pax5  $\beta$  gene is methylated in approximately 60% of breast and lung tumors and 20% of colon tumors.

Accordingly, the invention finds beneficial use, in association with the techniques disclosed in U.S. Patent No. 6,017,704 and PCT Application No. PCT/US01/26452, for the early detection of cancer, monitoring tumor progression, monitoring the response to radio- and chemotherapy, predicting disease outcome, and monitoring response to prevention therapy. The invention improves upon known cancer screening technology because it provides additional markers for early detection, monitoring of therapy, and prediction of disease outcome.

An initial step according to the invention was the identification of PAX5  $\alpha$  and  $\beta$  hypermethylation. The MCA/RDA technique developed by Toyota *et al.*, was used to

identify genes methylated in breast cancer (Toyota *et al.*, 1999). A subtractive library was constructed, and 100 clones were analyzed by DNA sequencing. Comparison of these clones revealed 50 unique sequences, and a Blast search of each clone confirmed that 48 of 50 clones were homologous to Genbank sequences located in the high-throughput genomic sequence database. One of these clones was identical to the 5' flanking region and exon 1  $\beta$  of the transcription factor PAX5 (Genbank accession # AF074913). Analysis of the genomic structure of this gene revealed the presence of two distinct promoters resulting in two alternative 5' exons (1 $\alpha$  and 1 $\beta$ ) that are spliced to common exons 2–10 resulting in the translation of two unique proteins due to a frameshift (see Busslinger *et al.*, 1996). Inspection of the  $\alpha$  and  $\beta$  promoters revealed that each region is representative of a CpG island (see Gardiner-Garden and Fromme, 1987; Antequera and Bird, 1993), with a GC content of 0.68 and 0.69; a CpG/GpC ratio of 0.70 and 0.73; and a total of 53 and 162 CpG sites in a 520 bp and 1,780 kb region, respectively.

A MSP analysis of the PAX5  $\alpha$  and  $\beta$  genes was then undertaken. Combined bisulfite restriction analysis (COBRA) analysis (8) was initially conducted on DNA from ten breast cell lines, five lung cancer cell lines, and five normal lymphocytes to screen for aberrant methylation of the PAX5  $\alpha$  and  $\beta$  genes. Following bisulfite modification, PCR products were produced using stage-1 primers that do not discriminate between methylated and unmethylated alleles. Methylated alleles were detected in the PAX5  $\alpha$  (389 bp) and  $\beta$  (328 bp) products following digestion with the restriction enzyme BSTU1, which specifically cleaves at CGCG sites that are retained following bisulfite modification due to the presence of methylated CpGs. Complete digestion of the PCR products indicative for methylation of the PAX5  $\alpha$  gene was observed in seven breast and five lung cancer cell lines. The PAX5  $\alpha$  PCR product was partially digested (10-80%) for two breast and lung cancer cell lines. Similarly, the PAX5  $\beta$  gene was completely digested in five breast and three lung cancer cell lines. The PAX5  $\beta$  PCR product was also partially digested (10-80%) for two breast and one lung cancer cell lines. PCR products from normal lymphocytes remained undigested, suggesting a lack of methylation (data not shown).

The frequency of PAX5  $\alpha$  and  $\beta$  methylation was then characterized in a panel of primary tumors (breast, lung, and colon) using applicants' two-stage MSP approach (see Patent Cooperation Treaty Application No. PCT/US01/26452, filed August 24, 2001).

5 The results are summarized in Fig. 1, and a representative MSP analysis is shown in Fig. 2 for both the PAX5  $\alpha$  and PAX5  $\beta$  genes. Fig. 2 depicts MSP analysis of the PAX5  $\alpha$  and  $\beta$  genes in a panel of primary breast tumors. Bisulfite-modified DNA was amplified with stage-2 primers specific for methylated alleles. The MDA-MB-231 and H2009 represent positive and negative cell lines for methylation of both the PAX5  $\alpha$  and  $\beta$  genes,  
10 respectively. In both breast and lung tumors, the frequency of PAX5  $\alpha$  methylation approximated 70% and tended to be greater than that observed for the PAX5  $\beta$  gene (54%). The frequency of PAX5  $\alpha$  and  $\beta$  methylation in colorectal tumors was significantly less than that in either lung or breast tumors ( $p < 0.0001$ ), but methylation of the  $\beta$  gene was strongly associated ( $p = 0.05$ ) with methylation of the  $\alpha$  gene in lung  
15 tumors and cell lines (Fig. 1).

MSP analysis for methylation in cell lines corroborated that seen by the COBRA assay. Overall, 90% and 60% of breast cancer-derived cell lines were methylated for the PAX5  $\alpha$  and  $\beta$  genes, respectively. Eleven lung tumor-derived cell lines were analyzed  
20 by MSP, and 82% and 55% were methylated for the PAX5  $\alpha$  and  $\beta$  genes, respectively. Methylated alleles were also detected at a lower prevalence (3–13%) in cultured nonmalignant bronchial epithelial cells (BEC) and sputum from cancer-free, high-risk subjects. Neither PAX gene was methylated by MSP in blood lymphocytes from nonsmokers.

25

Reverse transcription-PCR (RT-PCR) was conducted on cDNA from three breast cancer cell lines (MCF7, MDA-MB 231, and T47D) and one lung cancer cell line (Calu6) to determine the relationship between methylation of the PAX5  $\alpha$  and  $\beta$  genes and transcription. COBRA analysis demonstrated complete digestion for the PAX5  $\alpha$  gene in  
30 the MDA-MB-231 and T47D cell lines and partial digestion for MCF7 (20%) and Calu6



(80%). For the PAX5  $\beta$  gene, complete digestion was seen in the MCF7, MDA-MB 231, and Calu6 cell lines, but no digestion for T47D. Reference is made to Fig. 3A, showing an MSP analysis of breast and lung cancer cell lines. Bisulfite-modified DNA was amplified with stage-2 primers specific for methylated alleles. Fig. 3B shows an RT-PCR analysis for expression of the PAX5  $\alpha$  and  $\beta$  genes in breast and lung cancer cell lines, which were grown in the presence (+) and absence (-) of 1  $\mu$ M DAC for 72 h. Expression of the PAX5  $\alpha$  and  $\beta$  genes was restored in non-expressing cell lines following DAC treatment. Expression of the  $\beta$ -actin gene was determined as a control for RNA integrity. As indicated in Fig. 3A, these methylation changes were then corroborated by MSP.

10 PAX5  $\alpha$  transcript was abundant in MCF7 and to a lesser extent present in Calu6, while no transcript was present in the MDA-MB 231 and T47D cell lines, as seen in Fig. 3B. Treatment of these cell lines with 1  $\mu$ M DAC, an inhibitor of DNA methyltransferase, increased expression in Calu6 and restored expression in MDA-MB231 and T47D. PAX5  $\alpha$  transcript was only detected in T47D; however, treatment with DAC restored expression

15 in the other three cell lines. Treatment with DAC did not affect the expression of the housekeeping gene  $\beta$ -actin.

Methylation density within the PAX5  $\alpha$  and  $\beta$  promoter regions amplified by the MSP primers was then determined for the MCF-7 and T47D cell lines and three primary breast tumors. The PCR primers used for the COBRA analysis were also used to amplify the modified DNA from the cell lines, while methylation-specific primers were used to detect methylated sequences in the primary tumors where contaminating stromal and inflammatory cells are present. Sixteen and 13 CpG sites spanning the 116 and 86 bp regions between the MSP primers for the PAX5  $\alpha$  and  $\beta$  sequence, respectively, were

25 evaluated for comparing methylation density between cell lines and tumors. Reference is made to Figs. 3A, 3B, and 4A and 4B, illustrating how the sequencing results corroborated the expression studies, with 78% of sites methylated within the PAX5  $\alpha$  promoter in the T47D cell line, and 100% of sites methylated within the PAX5  $\beta$  promoter in the MCF-7 cell line. Figs. 4A and 4B graphically show the methylation density of the PAX5  $\alpha$  and  $\beta$

30 genes. Genomic DNA from breast cancer cell lines T47D and MCF-7, and three primary

breast tumors were treated with sodium bisulfite. PAX5  $\alpha$  and  $\beta$  PCR products containing 16 and 13 CpG sites, respectively, were generated. The PCR products were cloned, and five individual clones/sample were sequenced. Summary data are presented as the percent of CpG sites methylated. The CpG methylation density in all three primary  
5 tumors was very similar to the patterns observed in the methylated cell lines for each gene, consistent with these regions being important sites for methylation and silencing of the PAX5  $\alpha$  and  $\beta$  genes (See Fig. Figs. 4A and 4B).

The PAX gene family consists of nine members, each of which share a common  
10 motif called the paired box that displays DNA-binding properties (Stuart *et al.*, 1995). PAX proteins function as nuclear transcription factors important for cellular differentiation, migration, and proliferation (Schafer, 1998). That these genes are strong transcriptional regulators makes them likely targets for disruption in oncogenesis. Consistent with this premise, inappropriate expression of the PAX5 gene has been  
15 implicated in the pathogenesis of small lymphocytic lymphoma cancer and advanced stage glioblastoma (Schafer, 1998). According to the present invention, the PAX5  $\alpha$  and  $\beta$  genes are common targets for inactivation by aberrant promoter hypermethylation in lung, breast, and (to a lesser extent) colorectal tumors. Both genes exhibit homogeneous and dense methylation patterns that correlated with loss of transcription. Additionally  
20 according to the present invention, methylation of the PAX5  $\beta$  gene is associated with methylation of the  $\alpha$  gene in lung tumors, so there is a selective advantage to target both genes for inactivation in this cancer type.

Previous studies using MCA and RDA identified the PAX6 gene that is involved in  
25 eye, nose, pancreas, and brain development as a common target for aberrant methylation in colonic mucosa during aging (Toyota *et al.*, 1999a; Toyota *et al.*, 1999b). In these studies, the CpG island identified was localized within an enhancer present in the 5' region of the PAX6 gene. A CpG island within the exon 5 coding region of the PAX6 gene was also hypermethylated in bladder and colon tumors (Salem *et al.*, 2000). While  
30 methylation of the promoter region and exon 5 was common in cell lines, methylation in primary tumors was largely confined to exon 5 and did not affect gene transcription. This

is in marked contradistinction from the present invention using the PAX5  $\alpha$  and  $\beta$  genes. According to the present invention, dense methylation is seen in both promoter regions in derived cell lines and primary lung and breast cancers that correlated directly with loss of transcription. The difference seen in primary tumors for silencing the PAX5 and PAX6  
5 genes may be linked to their functions in the respective tissues. No expression of the PAX6 gene was seen in normal colonic mucosa, thereby negating any selective advantage for methylation of the promoter region; again in contradistinction, according to the present invention an abundant expression of both PAX5 transcripts in normal lung tissue is observed.

10

The fact that both PAX5 genes are expressed in normal lung suggests that these proteins are likely also modulating gene activity in pulmonary cells. The short arm of chromosome 9 is a frequent site for loss of heterozygosity in lung cancer, with deletions extending from 9p13–9p21 in approximately 50% of non-small cell lung cancers (*Testa et al.*, 1994). Both the PAX5 (9p13) and p16 (9p21) genes are localized to this region and  
15 now share the commonality of being inactivated frequently in lung cancer by aberrant promoter methylation (*Esteller et al.*, 2001; *Belinsky et al.*, 1998). There is also precedent for inactivating genes that code for transcription factor binding proteins by promoter hypermethylation. Hypermethylation in cancer (HIC-1) is a zinc-finger  
20 transcription factor gene that is commonly expressed in normal tissues, but inactivated by promoter hypermethylation in lung, breast, colon, and hematopoietic tumors (*Wale et al.*, 1995; *Issa et al.*, 1997). Similar to the PAX5 gene, HIC1 is also important in development with knockout mice dying perinatally and exhibiting gross developmental defects involving the brain, cleft palate, and limbs (*Carter et al.*, 2000).

25

Changes in gene-specific methylation may serve as intermediate biomarkers for cancer detection, risk assessment, and monitoring disease in sputum and blood (*Palmisano et al.*, 2000; *Esteller et al.*, 1999). For example, respecting lung cancer, the aberrant methylation of the p16 and/or O<sup>6</sup>-methylguanine-DNA methyltransferase  
30 promoters was detected in DNA from sputum of patients with squamous cell carcinoma (SCC) up to three years before clinical diagnosis (*Palmisano et al.*, 2000).

According to the present invention, the PAX5  $\alpha$  and  $\beta$  genes constitute new intermediate markers for evaluation. Although the timing for inactivation of these genes is not precisely defined, the low frequency of methylation seen in bronchial epithelium and sputum from cancer-free, high-risk subjects indicates that these genes may be methylated during progression rather than initiation. This contrasts with applicant's finding with p16, which is inactivated at the earliest cytologic stages of SCC and adenocarcinoma (ADC), a fact that may account for its more common detection (35%) in bronchial epithelium and sputum from current and former smokers. See Patent Cooperation Treaty Application No. PCT/US01/26452, filed August 24, 2001, assigned to the assignee of the present invention. The utility of plasma or serum for detecting circulating aberrantly methylated DNA in patients with colorectal, head and neck, and breast cancers is known (**Zou *et al.*, 2002, Silva *et al.*, 2002**). Thus, the inclusion according to the present invention of the PAX5 genes into molecular marker panels for lung, breast, and colon cancer improves the sensitivity and specificity for developing risk models for detecting these cancers through analysis of sputum and blood in high-risk subjects.

Thus the PAX5  $\alpha$  and  $\beta$  genes that are inactivated by promoter hypermethylation demonstrate utility as biomarkers for predicting cancer risk. A panel of genes, inactivated by promoter hypermethylation, has been identified as biomarkers for predicting lung cancer risk and for early detection. Identification occurred in a nested, case-control study conducted on subjects recruited through the University of Colorado. The establishment of a cohort of current and former smokers targeting persons with chronic obstructive pulmonary disease (COPD) and smoking histories in excess of 30 pack-years was initiated. Home sputum was collected longitudinally (optimally once a year) on all enrolled subjects. Enrollment in this cohort has now reached 3,000 persons, and approximately 120 incident cases of lung cancer have been observed.

In an initial study, 33 incident cases of lung cancer were selected and matched 1:1 by age, gender, and date of first sputum sample to 33 persons who were still clinically cancer free (controls). Sputum samples collected on controls after the matched case was diagnosed with cancer were censored. Methylation of the PAX5 alpha and beta represented two of the genes assessed in sputum specimens in the laboratory with blinded

replicates for quality control and coded identifiers for all subjects such that our laboratory has remained blinded as to case-control status. The number of sputum samples collected per person ranged from one to five specimens. PAX5 alpha and beta methylation in sputum was associated with odds ratio of 2.0 and 2.4, respectively. Thus, these findings suggest that methylation of the PAX5 genes is an important component of a panel of markers for predicting lung cancer risk.

Through a gene discovery program, several other candidates possessing CpG rich promoter regions have emerged as gene targets in cancer. Three CpG islands have been identified, two of which predict for genes, while the other remains to be characterized. The characteristics of each island are shown below.

	<u>CpG Island</u>	<u>%GC Content</u>	<u>Chromosome Location</u>
15	1. Novel helix loop helix binding protein homologous to the murine gene BHLHB4	72%	20q13
	2. Novel gene	69%	20q12-13.2
20	3. Beta 3	72%	8q

The first two CpG islands are clustered within chromosomal region 20q13. The novel helix loop helix gene is homologous to BHLHB4, a gene isolated from a mouse pancreatic  $\beta$ -cell line. The murine gene is proposed to modulate the expression of genes required for the differentiation and/or maintenance of pancreatic and neuronal cell types. The Beta 3 gene is a single exon gene that also codes for a basic helix loop helix protein. Applicants' studies to date indicate that all three CpG islands are inactivated by promoter hypermethylation in approximately 80% of breast cancer cell lines evaluated ( $n = 8-10$ ). In contrast, CpG island 1 is commonly inactivated in lung cancer cell lines (8 of 11) while the prevalence for inactivation of the other two CpG islands in cell lines is <10%.

Applicants have extended studies of CpG island 1 to primary squamous cell carcinoma of the lung and find inactivation by methylation in 60% of tumors evaluated. Thus, these 3 novel genes likely may be involved in the development of breast and/or lung cancer and could be valuable as biomarkers for early detection, prognosis, and monitoring

the efficacy of preventive interventions. In addition, these genes may also prove in future studies to be inactivated in other solid and liquid tumors (e.g., leukemia), making them equally valuable as biomarkers for other cancers. The following are the primer sequences and annealing temperatures for the Novel Helix Loop Helix Binding Protein, Novel

Gene 2, and Beta 3 Genes.

#### Stage 1 PCR

<u>Gene</u>	<u>Primers (5' - 3')</u>	<u>Annealing Tm</u>	<u>Product Size</u>
HLHBP	gaggagaggaggtgggagag (SEQ ID NO:9) Crtaaccrtaacttaataccaaatac (SEQ ID NO:10)	58	267 bp
Novel Gene 2	gtttagttyggaggaaggattttta (SEQ ID NO:11) Taataataatccaaatacrccaaacc (SEQ ID NO:12)	60	331 bp
Beta 3	aaagaaagaaggggagagggtttt (SEQ ID NO:13) Acaacaacaaccctaccccctc (SEQ ID NO:14)	60	393 bp

$$Y = C + T, R = A + G$$

#### Stage 2 PCR

<u>Gene</u>	<u>Primers (5' - 3')</u>	<u>Annealing Tm</u>	<u>Product Size</u>
HLHP	gaggaggtagcgggcgc (SEQ ID NO:15) Tcgaccataaccgcgcgc (SEQ ID NO:16)	66	186 bp
Novel Gene 2	ggtcggaataatagcgcgc (SEQ ID NO:17) Gaacgtccataacgaacgcgc (SEQ ID NO:18)	68	181 bp
Beta 3	tagtattaggatcgacgcgc (SEQ ID NO:19) Gtcctcgccgacgaccg (SEQ ID NO:20)	68	179 bp

Fig. 5 tabulates the frequency of methylation of the novel Helix Loop Helix Binding Protein, Novel Gene 2, and Beta3 genes in tumor cell lines, primary tumors and non-malignant specimens.

Industrial Applicability:

The invention is further illustrated by the following non-limiting example.

5       **Tissue Samples and Cell Lines.** Lung SCCs were obtained from patients previously enrolled in a Lung Cancer Surveillance Study conducted through St. Mary's Hospital, Grand Junction, Colorado. Lung ADCs were acquired from the Johns Hopkins Lung Spore Repository, Baltimore, Maryland. Breast tumors were collected from women enrolled in a New Mexico Women's Health Study being conducted within the  
10   Epidemiology and Cancer Control Program at the University of New Mexico. Colorectal carcinoma samples were collected from patients undergoing surgery at the Hospital de Sant Pau and the Hospital de Bellvitge in Barcelona and Catalonia, Spain, respectively. Sputum, nonmalignant BECs, and blood lymphocytes were obtained from veterans who use the Multispecialty Chest Clinic at the New Mexico Veterans Health Care System for  
15   their primary care. Sputum cytology was not diagnostic for cancer in any of these individuals and ranged from normal to marked dysplasia in this population. All subjects gave informed consent according to institutional guidelines. Tumor-derived cell lines were obtained from the American Type Culture Collection and were cultured according to their conditions.

20

**MCA/RDA.** MCA/RDA was performed as described by Toyota *et al.* (Toyota *et al.*, 1999a). Briefly, 5 µg of DNA from the cell line MCF7 was used as the tester, and a mixture of DNA from normal breast tissue of five women (1 µg each) was used as the driver. MCA amplicons were produced using RMCA adaptors, and two rounds of  
25   competitive hybridizations were performed. The resulting RDA products were digested with the restriction endonuclease *XmaI* (New England Biolabs) and subsequently cloned into pBluescript KS+.

**DNA Sequencing and Analysis.** Plasmid DNA containing the RDA products was  
30   prepared using the QIAprep Spin Miniprep according to the manufacturer's instructions (Qiagen). Virco (Cambridge, UK) analyzed the sequences using an automated DNA sequencer (Applied Biosystems). Sequence homologies were determined using the Blast

program of the National Center for Biotechnology Information  
(<http://www.ncbi.nlm.nih.gov/BLAST/>).

- MSP and Bisulfite Sequencing.** DNA was isolated by standard phenol-chloroform extraction and ethanol precipitation. Genomic DNA was modified by treatment with sodium bisulfite that converts only unmethylated cytosines to uracil. The methylation status of the PAX5  $\alpha$  and  $\beta$  genes was determined using a nested, two-stage method described previously (Palmisano *et al.*, 2000). Primer sequences used in the stage-1 amplification of each gene are as follows:
- 10         $\alpha$ -Forward, (5'-gggt ttgtatatggagatgttatagg-3') (SEQ ID NO:1);  
           $\alpha$ -Reverse, (5'-caacatcacaaaatatcccaaacac-3') (SEQ ID NO:2);  
           $\beta$ -Forward, (5'-agtttgtgggtgttagttaatgg-3') (SEQ ID NO:3); and  
           $\beta$ -Reverse, (5'-caaaaaatcccaaccaccaaacc-3') (SEQ ID NO:4).
- 15        All PCR amplifications were performed using a Biometra T3 thermocycler and Taq Gold polymerase (Perkin-Elmer). The cycling parameters for stage-1  $\alpha$  were as follows: 94°C for 10 min; then 40 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min; and a 5-min final extension at 72°C. Stage-1  $\beta$  conditions were identical except the annealing temperature was reduced to 54°C, and all cycling times were performed for
- 20        30 s. The size of the stage-1  $\alpha$  and  $\beta$  PCR products was 389 and 328 bp, respectively. Primer sequences used in the stage-2 amplification of each gene are as follows:
- $\alpha$ -BSM1, (5'-ataaaagtttggggcggcgc-3') (SEQ ID NO:5);  
           $\alpha$ -BSM2, (5'-gcgccccaacgcgccg-3') (SEQ ID NO:6);  
25         $\beta$ -BSM1, (5'-gagttgagtttcgggcggc-3') (SEQ ID NO:7); and  
           $\beta$ -BSM3, (5'-gccgccgcccgcgcg-3') (SEQ ID NO:8).

The cycling parameters for stage-2  $\alpha$  were as follows: 94°C for 10 min; then 40 cycles of 94°C for 15 s, 66°C for 15 s, and 72°C for 15 s; and a 5-min final extension at

30        72°C. Stage-2  $\beta$  conditions were identical except the annealing temperature was decreased to 64°C. The size of the stage 2- $\alpha$  and  $\beta$  PCR products was 166 and 124 bp,



respectively.

DNA isolated from cell lines MDA-MB-231 and NIH-2009 served as positive and negative controls, respectively, for both genes. All assays were conducted in at least  
5 duplicate and were confirmed positive for methylation by restriction digestion with *Bst*UI or by DNA sequencing.

Bisulfite-modified DNA from the T47D and MCF-7 cell lines was amplified using the stage-1 primers for the PAX5  $\alpha$  and  $\beta$  genes. Stage-1 and methylation-specific stage-  
10 2 primers were used to amplify modified DNA from three breast tumors shown to be methylated for the  $\alpha$  or  $\beta$  gene. The PCR products were ligated into the PCR II vector using the TA cloning kit (Invitrogen, San Diego, California). Five clones from each sample were sequenced.

15 **5-Aza-2'-deoxycytidine Treatment and RT-PCR.** Re-expression studies were performed using three breast cancer cell lines (MCF7, MDA-MB-231, and T47D) and one lung cancer cell line (Calu6). Cell lines were treated for 3 days in culture media with 1  $\mu$ M DAC (Sigma Chemical Co.). Total RNA was prepared using Trizol (Life Technologies, Inc.), and 3  $\mu$ g aliquots were reverse-transcribed using the Superscript kit  
20 (Life Technologies, Inc.). The expression of the PAX5  $\alpha$  and  $\beta$  transcripts were determined by reverse transcription (RT) -PCR using the exon 1 $\alpha$  forward primer (5'-cctgtccattccatcaagtctg-3') and the exon 1 $\beta$  forward primer (5'-cccgatggaaatacactgtaagcac-3') with an exon 2 reverse primer (5'-ttttgctgacacaacctggctgac-3').  $\beta$ -actin was also amplified as a control for RNA  
25 integrity.

PCR amplification was performed at 94°C for 10 min; then 35 cycles of 94°C for 30 s, 64°C for 30 s, and 72°C for 30 s; and a 5-min final extension at 72°C. PCR products were analyzed on a 3% agarose gel containing ethidium bromide and visualized under UV  
30 illumination.

The proportion of tumors positive for methylation of PAX5  $\alpha$  or  $\beta$  was compared among cancer types with the Fisher's exact test. The association between PAX5  $\alpha$  and  $\beta$  methylation for each cancer type was also assessed by the Fisher's exact test.

- 5        The preceding example can be repeated with similar success by substituting the generically or specifically described reactants and/or operating conditions of this invention for those used in the preceding examples.

- 10        Although the invention has been described in detail with particular reference to these preferred embodiments, other embodiments can achieve the same results. Variations and modifications of the present invention will be obvious to those skilled in the art and it is intended to cover in the appended claims all such modifications and equivalents. The entire disclosures of all references, applications, patents, and publications cited above are hereby incorporated by reference.

15